

Nut en noodzaak van kerngegevens in StUF.

Roel de Bruin, 24-11-2016

De Stuf- standaard beschrijft de kerngegevens als volgt (stuf0301.pdf, blz 23):

Om problemen met de identificatie te vermijden kent StUF het begrip kerngegevens. Dit is een deelverzameling van de attributen en relaties van een entiteitstype aan de hand waarvan een object kan worden geïdentificeerd. Bij personen kan bijvoorbeeld voor de volgende set identificerende gegevens gekozen worden:

- *Burgerservicenummer*
- *A-nummer*
- *Naamgegevens (Geslachtsnaam, voorvoegsel en voorletters)*
- *Verblijfsadres*
- *Geboortedatum*
- *Geslacht*

Uit bovenstaande blijkt dat de kerngegevens voor entiteitstypen worden gedefinieerd met als doel het identificeren van objecten. Op zich is dat een goed idee maar mijns inziens spelen hierbij de volgende problemen:

1. De kerngegevens zijn niet expliciet vastgelegd.

Op dit moment zijn de kerngegevens alleen vastgelegd in specifieke complextypes in de xsd-schema's met een naam die eindigt op 'kerngegevens'. Zij worden bepaald in het proces van verstufing van het informatiemodel, maar zijn niet in dat model gespecificeerd.

Ik denk dat, als we kerngegevens blijven gebruiken, het verstandig zou zijn deze expliciet in het model te specificeren.

2. Er is niet gedefinieerd welke combinaties van kerngegevens identificierend zijn.

In de StUF specificaties wordt gesteld dat de kerngegevens verplicht moeten worden opgenomen in een aantal berichttypen (en onder bepaalde omstandigheden). Daarbij wordt voorbijgegaan aan het feit dat niet alle kerngegevens nodig zijn om een voorkomen uniek te identificeren. Waarschijnlijk is dit gedaan om de standaard niet te complex te maken, maar het leidt in het gebruik ervan wel tot onnodig ingewikkelde berichten. Immers, als het BSN van een persoon bekend is, is de rest van de lijst van kerngegevens niet nodig voor identificatie.

3. Er is niet gedefinieerd of een kerngegeven verplicht is.

Er wordt in de specificatie niet over gerept of een kerngegeven al dan niet een geldige waarde moet bevatten. In de schema's zijn alle gegevens optioneel en nillable. Dat betekent dat een kennisgevingsbericht dat alle kerngegevens bevat zonder waarde ook voldoet aan standaard. Bovendien betekent dit ook dat in de praktijk de opsteller van een vraagbericht er niet vanuit kan gaan dat een subset van kerngegevens die op zich identificerend is ook het verwachte resultaat zal opleveren. En zou de opsteller van het vraagbericht voor de zekerheid alle kerngegevens opnemen als condities, dan is er nog geen garantie dat het verwachte antwoord wordt gevonden, want als de ontvanger van de vraag één van die kerngegevens niet kent, dan voldoet het betreffende object niet aan alle gestelde voorwaarden.

Ik denk dat het niet doenlijk is alle kerngegevens verplicht te maken in die zin dat er ook verplicht een geldige waarde moet worden opgegeven, maar uit bovenstaande blijkt wel dat de kerngegevens in de huidige vorm niet ondersteunend zijn in de vraag- en antwoord-berichten. Waar in het verleden de nadruk lag op de uitwisseling van data zien we nu een tendens naar centrale opslag en een toenemend gebruik van vraag- en antwoord-berichten. Ik denk dat het dus verstandig is te zoeken naar een nieuwe vorm van identificatie van voorkomens.

Op bladzijde 23 van het document stuf0301.pdf staat ook :

Zeker betekenisloze sleutels zullen niet voorkomen in het sectormodel, waar per slot van rekening de werkelijkheid wordt gemodelleerd en niet de opslag van gegevens in een database. Omdat deze sleutels van belang zijn bij de identificatie van voorkomens, worden deze sleutels in StUF als afzonderlijk attribuut onderkend.

In de tijd dat deze tekst werd opgesteld was dit inderdaad een correcte constatering, maar deze is door de tijd ingehaald. De informatiemodellen RSGB en RGBZ bevatten tegenwoordig een groot aantal betekenisloze sleutels (identificatie). Ik denk dat het niet nodig is voor entiteitstypen die een dergelijke sleutel kennen, andere kerngegevens te specificeren.

Bovendien lijkt het mij ook niet meer nodig uitwisseling op basis van systeemsleutels te ondersteunen voor entiteiten die een dergelijke betekenisloze sleutel bevatten. In ieder geval niet voor de recente standaarden zoals RGBZ.

In de praktijk blijkt ook dat gedefinieerde koppelvlakken, zoals bijvoorbeeld de Prefill eFormulierservices en de Zaak- en Documentservices, de in het achterliggende halffabricaat gedefinieerde kerngegevens negeren. Alleen de identificatie-attributen worden gebruikt voor unieke identificatie van de objecten.

Het grote voordeel van deze identificatie attributen is dat we daarmee de kerngegevens hebben teruggebracht tot één attribuut per entiteitstype waardoor het geen probleem is dit verplicht te stellen. Objecten zijn daarmee altijd uniek identificeerbaar, in kennisgevingen van het type 'wijziging' hoeven geen overbodige gegevens meer te worden opgenomen en de beschreven problemen in de vraag- en antwoord berichten zijn daarmee opgelost.

Vaak worden de kerngegevens ook gezien als de meest relevante eigenschappen van een entiteitstype die in ieder geval uitgewisseld moeten worden. Aangezien in de huidige xsd-schema's de kerngegevens zowel optioneel als nillable zijn, is dit aspect onvoldoende uitgewerkt. Het zou al een verbetering zijn als in de schema's onderscheid wordt gemaakt tussen toevoegingen en wijzigingen. Daarmee kan in de definities van de toevoegberichten per element expliciet worden aangegeven of het verplicht moet worden opgenomen en of het al dan niet leeg mag worden gelaten.

Met de afspraak om in de toekomst de bestaande sectormodellen te vervangen door koppelvlakken wordt deze aanpak alleen maar eenvoudiger.